AD___________________

Award Number: DAMD17-02-1-0558

TITLE: Statistical Inference for Quality-Adjusted Survival Time

PRINCIPAL INVESTIGATOR: Hongwei Zhao, Sc.D.

CONTRACTING ORGANIZATION: University of Rochester
Rochester, NY 14642

REPORT DATE: July 2006

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*

**1. REPORT DATE** *(DD-MM-YYYY)*
01-07-2006

**2. REPORT TYPE**
Annual Summary

**3. DATES COVERED** *(From - To)*
15 JUL 2002 - 14 JUN 2006

**4. TITLE AND SUBTITLE**
Statistical Inference for Quality-Adjusted Survival Time

**5a. CONTRACT NUMBER**

**5b. GRANT NUMBER**
DAMD17-02-1-0558

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**
Hongwei Zhao, Sc.D.

E-Mail: Hongwei_Zhao@urmc.rochester.edu

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

University of Rochester
Rochester, NY 14642

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

**10. SPONSOR/MONITOR'S ACRONYM(S)**

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

**12. DISTRIBUTION / AVAILABILITY STATEMENT**
Approved for Public Release; Distribution Unlimited

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT:**

In evaluations of breast cancer therapies, the patients' quality of life is receiving more and more attention. It is desirable that a treatment not only prolongs the overall survival life, but also improves the quality of life (QOL). Quality-adjusted lifetime (QAL) is a measure that combines both the quality and the quantity of a person's lifetime. The goal of my research is to study how to draw inference about QAL in the presence of censoring. For the grant period, I have studied the following problems: (1) Estimating survival functions of QAL (2) Testing the equality of two survival functions of QAL. (3) Testing the equality of survival functions of QAL from three or more groups. (4) Developing regression methods for evaluating the effects of covariates on QAL. My research has resulted in a manuscript that has been submitted and invited talks at statistical conferences.

**15. SUBJECT TERMS**
Quality of life, censored data, hypothesis testing, regression, survival analysis.

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON<br>USAMRMC |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | 19b. TELEPHONE NUMBER *(include area code)* |
| U | U | U | UU | 38 | |

## Table of Contents

# Annual Summary Report

Statistical Inference for Quality-Adjusted Survival Time

Hongwei Zhao, Sc.D

## a. Introduction

In studies of treatment options for breast cancers, it is desirable to find a treatment that not only prolongs the overall survival life, but also improves the quality of life (QOL). Quality-adjusted lifetime (QAL) is a measure that combines both the quality and the quantity of a person's lifetime. First proposed by Gelber, Gelman and Goldhirsch (1989), QAL is simply an integration of survival time weighted by a utility coefficient ranging from 0 (poor health) to 1 (perfect health). In a typical clinical trial setting, patients are enrolled over time, and the study ends before observation of the endpoints for all patients. Therefore, the data are right censored. The goal of my research is to study how to draw inference about QAL in the presence of censoring.

## b. Body

During the grant period from July 2002 to July 2006, I studied the problem of making statistical inference on the quality adjusted lifetime, when censoring is present. These problems include (1) Estimating survival functions of QAL (2) Testing the equality of two survival functions of QAL. (3) Testing the equality of survival functions of QAL from three or more groups. (4) Developing regression methods for evaluating the effects of covariates on QAL.

### b.1. The Setting

For the $i$th individual in the study, let's define $V_i(t), t \geq 0$, as a continuous time stochastic process representing the patient's health history process, $T_i$ as the survival time. $U$ is a utility function, which is assumed to be known or can be specified. Denote ${V_i}^H(t)$ as the health history information up to time $t$, i.e. ${V_i}^H(t) = \{V_i(u) : u \leq t\}$. The $i$th individual's quality adjusted lifetime, denoted as $Q_i$, is equal to

$$Q_i = \int_0^{T_i} U\{V_i(t)\}dt.$$

Let $Z_i$ denote the $p+1$ vector of covariates associated with the $i$th individual, and $C_i$ denote the censoring variable. We assume that the censoring variable is independent of the health history, conditional on the covariates.

Due to the presence of censoring, we cannot make inference on QAL over the entire health history. We can only consider the QAL accumulated within a time L, where

L is a time limit up to which we have reasonable amount of data. Our inference therefore will depend on the choice of L. Consequently, the survival time of a patient will be truncated at L, that is, $T^L = \min(T, L)$. For ease of notation, however, we still use $T$ instead of $T^L$ in subsequent development of the theory.

We observe the following data:

$$X_i = \min(T_i, C_i), \Delta_i = I(T_i < C_i), V_i(t), 0 \le t \le X_i, Z_i,$$

$$Q_i = \int_0^{X_i} U\{V_i(t)\}dt, \quad i = 1, \cdots, n.$$

We would like to make inference about the true distribution of $Q_i$, and investigate which covariates affect its distribution.

### b.2. Estimating survival functions of quality-adjusted lifetime

A simple weighted estimator for the survival function of $Q_i$, $S(x) = \Pr(Q_i > x)$, can be formed by

$$\widehat{S}_{WT}(x) = n^{-1} \sum_{i=1}^{n} \frac{\Delta_i B_i}{\widehat{K}(T_i)},$$

where $B_i = I(Q_i > x)$, $\widehat{K}(T_i)$ is the Kaplan-Meier estimator of the survival function for the censoring variable $C$, $K(T_i) = \Pr(C > T_i)$.

Zhao and Tsiatis (1997, 1999) outlined the form for the most efficient estimator for the survival function of QAL:

$$\widehat{S}_{eff}(x) = n^{-1} \sum_{i=1}^{n} \frac{\Delta_i B_i}{\widehat{K}(T_i)} + n^{-1} \sum_{i=1}^{n} \int_0^{\infty} \frac{dM_i^c(u)}{K(u)} [e_{eff}\{V_i^H(u)\} - G(e_{eff}, u)]$$

where $\mathcal{M}_i^C(u) = N_i^C(u) - \int_0^u \lambda^C(t) Y_i(t) dt$, $N_i^C(u) = I(X_i \le u, \Delta_i = 0)$, $Y_i(t) = I(X_i \ge t)$, $\lambda^C(t) = \lim_{h \to 0} \frac{1}{h} \Pr(t < C < t + h | C \ge t, T \ge t)$ is the hazard function for the censoring distribution. $e_{eff}\{V_i^H(u)\} = \mathrm{E}\{B_i | V_i^H(u)\}$. For any random variable $X$, the function $G(X, u)$ is defined as

$$G(X, u) = \frac{\mathrm{E}\{X_i I(T_i \ge u)\}}{S_T(u)}.$$

Here, $S_T(u) = \mathrm{pr}(T > u)$.

Since $e_{eff}\{V_i^H(u)\}$ cannot be estimated non-parametrically, we cannot obtain the most efficient estimator. Instead, our goal is to find an improved estimator, which is more efficient than the simple weighted estimator:

An improved estimator has been proposed by Zhao and Tsiatis (1999) which has the following form:

$$\widehat{S}_{imp}(x) = n^{-1}\sum_{i=1}^{n}\frac{\Delta_i B_i}{\widehat{K}(T_i)} + n^{-1}C\sum_{i=1}^{n}\int_0^\infty \frac{dM_i^c(u)}{K(u)}[e\{V_i^H(u)\} - G(e,u)],$$

where

$$e\{V_i^H(u)\} = Q_i(u),$$

and

$$C = \frac{\operatorname{cov}\left[\int_0^\infty \frac{dM_i^c(u)}{K(u)}\{B_i - G(B,u)\}, \int_0^\infty \frac{dM_i^c(u)}{K(u)}[e\{V_i^H(u)\} - G(e,u)]\right]}{\operatorname{var}\left[\int_0^\infty \frac{dM_i^c(u)}{K(u)}[e\{V_i^H(u)\} - G(e,u)]\right]}.$$

This estimator is asymptotically always more efficient than the simple weighted estimator. However, with finite sample size, it may not always perform well due to the fact that the coefficient $C$ has to be estimated. With simulation studies, we have found that an improved estimator that is more reliable than this estimator is one that uses $C = 1$. The advantages of this estimator include having a smaller variance with a small sample size, being easier to calculate, and having a more accurate variance estimator.

### b.3. Testing equality of survival functions of quality-adjusted lifetime

If an influence function for a test statistic exists for complete data case, denoted as $\psi_i$, then a test statistic for censored case can be constructed as

$$n^{-1}\sum_{i=1}^{n}\frac{\Delta_i \psi_i}{\widehat{K}(T_i)}, \tag{1}$$

where $\Delta_i$ is an indicator whether the subject $i$'s death is observed, and $\widehat{K}(T_i)$ is an estimator for the survival function for the censoring variable.

Zhao and Tsiatis (2001) proposed a test statistic where $\psi_i$ is the influence function of the general logrank test:

$$\psi_i = \int_0^\infty w(u)\left[Z_i - \frac{E\{Z_i I(Q_i \geq u)\}}{E\{I(Q_i \geq u)\}}\right] dM_i^Q(u),$$

where $Z_i$ is the treatment indicator, $Q_i$ is the quality adjusted survival time for subject $i$, $w(u)$ is any weighting function, $M_i^Q(u) = I(Q_i \leq u) - \int_0^u \lambda^Q(t) I(Q_i \geq t)dt$, $\lambda^Q(t)$ is the hazard function for $Q_i$, which under the null hypothesis is independent of $Z_i$. They showed that by choosing a certain weight $w(u)$, the test statistic became the ordinary logrank test when the utility coefficient is equal to 1 everywhere until a subject's death.

We have considered other forms of test statistics. One option is to use Pepe and Fleming (1989) test with censored data, where the survival function for each treatment can be estimated consistently using methods of Zhao and Tsiatis (1997). Specifically, the statistic for the difference between integrated quality of life adjusted survival curves for two treatment groups is:

$$\sqrt{n}\int_0^{\tau}\{\widehat{S}_1(x)-\widehat{S}_2(x)\}dx. \tag{2}$$

We found that the relative efficiency of these two test statistics depends highly on the shape of the survival curves of the QAL and neither one is dominant of the other.

**b.4. Testing equality of survival functions of quality-adjusted lifetime for three or more groups**

Both of these tests (**??**) and (**??**) can be extended to the case of testing the equality of QAL from $K$ groups ($K > 2$). Denote $Z_k$, $k = 1, \cdots, K-1$, to be the test statistics for $K-1$ two-group comparisons, and the estimated variance-covariance matrix of these statistics is denoted by matrix $\Sigma$. Then a test statistic for testing the equality of the K samples can be formed by

$$(Z_1, \cdots, Z_{K-1})\Sigma^{-1}(Z_1, \cdots, Z_{K-1})^T.$$

This test statistic has a chi-squared distribution under the null hypothesis.

Simulations studies showed that the large sample theory works well with finite sample sizes.

**b.5. Regression models for the mean QAL**

We assume that the mean QAL depends on the covariates in a very general form:

$$E(Q_i|Z_i) = g(\beta, Z_i), \tag{3}$$

where $\beta$ is a $(p+1)\times 1$ vector of parameters of interest, and $g(.)$ is a known function. Special cases include $g(\beta, Z_i) = \beta' Z_i$, a linear regression model and $g(\beta, Z_i) = g(\beta' Z_i)$, a generalized linear regression model.

If complete data are observed, a consistent estimator $\hat{\beta}$ for $\beta$ in the mean model (**??**) can be obtained from the following estimating equation:

$$S_n^F(\beta) = \sum_{i=1}^{n} h(Z_i)\{Q_i - g(\beta, Z_i)\} = 0, \tag{4}$$

where $h(Z_i)$ is $(p+1)$-dimensional vector of functions of $Z_i$, and the superscript $F$ represents models for full data. From the semi-parametric theory (Robins and Rotnitzky, 1992), we know that the most efficient estimating equation for complete data case is the one with

$$h^F_{eff}(Z_i) = \mathrm{Var}(Q_i|Z_i)^{-1}\frac{\partial g(\beta, Z_i)}{\partial \beta}|\ _{\beta_0},$$

where $\beta_0$ is the true value of the parameters.

When censoring is present, $Q_i$ cannot be observed for everybody so the estimating equation (**??**) cannot be used. However, using the idea of inverse probability weighting, which was originally proposed by Horvitz and Thompson (1952), we can construct a simple weighted estimating equation for $\beta$ in our mean model (**??**) with censored data:

$$S_{n,WT}(\beta) = \sum_{i=1}^{n} \frac{\Delta_i}{\widehat{K}(T_i)} h(Z_i)\{Q_i - g(\beta, Z_i)\} = 0, \tag{5}$$

It is easy to show that this estimating equation will produce consistent estimators for $\beta$.

In the above estimating equation, only the data on QAL for the people who have failures are used, the QAL for censored subjects are ignored. Hence the above estimating equation is not efficient.

From the semi-parametric theory for missing data processes developed by Robins and Rotnitzky (1992) and Robins et al. (1994), the influence function for the estimating equation for any regular asymptotic linear (RAL) estimators of $\beta_0$ can be written as

$$D_i^h - \int_0^\infty \{D_i^h - G(D^h, u)\}\frac{d\mathcal{M}_i^C(u)}{K(u)} + \int_0^\infty \left[e\{V_i^H(u)\} - G[e\{V^H(u)\}, u]\right]\frac{d\mathcal{M}_i^C(u)}{K(\ u)} \tag{6}$$

where $D_i^h \equiv h(Z_i)\{Q_i - g(\beta, Z_i)\}$ is the influence function for the complete data, $e\{V_i^H(u)\}$ is any $(p+1)$-dimensional vector of functionals of the health history $V_i^H(u)$. It should be noted that the influence function for the simple weighted estimating equation (**??**) is simply the first two terms of (**??**) (Zhao and Tsiatis, 1997, equation A.7).

From Robins and Rotnitzky (1992), the most efficient estimating equation is obtained by choosing

$$e_{eff}\{V_i^H(u)\} = \mathrm{E}\{D_i^{h_{eff}}|V_i^H(u)\} = h_{eff}(Z_i)\mathrm{E}\{D_i|V_i^H(u)\},$$

where $D_i \equiv Q_i - g(\beta, Z_i)$, and

$$h_{eff}(Z_i) \quad = \quad \{\mathrm{Var}(Q_i|Z_i) + P(Z_i)\}^{-1}\frac{\partial}{\partial \beta} g(\beta, Z_i)|_{\beta_0},$$

with

$$P(Z_i) = \mathrm{E}\left[\int_0^\infty \frac{dN_i^C(u)}{K(u)} \mathrm{Var}\{D_i | V_i^H(u), Y_i(u) = 1, Z_i\} | Z_i\right] \tag{7}$$

From the above results, the most efficient estimating equation can be formed by

$$S_{n,eff}(\beta) = \sum_{i=1}^n \frac{\Delta_i D_i h_{eff}(Z_i)}{\hat{K}(T_i)} + \sum_{i=1}^n \int_0^\infty \{e_{eff} - \hat{G}^\star(e_{eff}, u)\} \frac{dN_i^C(u)}{K(u)} = 0, \tag{8}$$

where $\hat{G}^\star(W, u) = \sum_{i=1}^n W_i Y_i(u)/Y(u)$ is a consistent estimator for $G(W, u)$, for any functional W.

Due to the difficulty in obtaining the most efficient estimating equation nonparametrically, we wish to find an estimating equation which can be obtained from the observed data, and which can be more efficient than the simple weighted estimating equation for any choice of $h(Z)$.

We first considered a method for obtaining the improved estimating equation, which is similar to the approach appeared in Zhao and Tsiatis (1999) for obtaining improved estimators of mean QAL, and in Bang and Tsiatis (2002) for median regression of medical costs. We called this approach the Best-Coefficient approach. Our second strategy is to use $Q_i(u)$ in place of $\mathrm{E}\{Q_i | V_i^H(u)\}$ in the formula for $e_{eff}\{V_i^H(u)\}$, i.e. we choose

$$e\{V_i^H(u)\} \quad = \quad h(Z_i)\{Q_i(u) - g(\beta, Z_i)\} \equiv D_i^h(u).$$

The corresponding estimating equation, named the improved estimating equation, has the following form:

$$S_n^{IMP}(\beta) = \sum_{i=1}^n \frac{\Delta_i}{\hat{K}(T_i)} D_i^h + \sum_{i=1}^n \int_0^\infty \frac{dN_i^C(u)}{\hat{K}(u)} [D_i^h(u) - \hat{G}^\star\{D^h(\ u), u\}] = 0. \tag{9}$$

Our last strategy for improving efficiency is to estimate $\mathrm{E}\{Q_i | V_i^H(u)\}$ by regressing $\frac{\Delta_i Q_i K(u)}{K(T_i)}$ on observed covariates, using only those observations with $X_i \geq u$ (Robins and Rotnitzky, 1992). The resulting estimating equation will be in the same form as the improved estimating equation except that $Q_i(u)$ is replaced by the estimate of $\mathrm{E}\{Q_i | V_i^H(u)\}$ from the regression approach.

Extended simulation studies were carried out examining the efficiency of these different approaches under different simulation scenarios. We concluded that the improved estimator performs the best among these different approaches. The details are shown in the attached manuscript.

## c. Key Research Accomplishment

1. I have gained knowledge on how to obtain the utility coefficients for quality adjusted survival time.

2. I have a better understanding of the general representation theory for missing data process.

3. I have examined different estimators for the survival functions of QAL.

4. I have compared different test statistics for testing equality of survival functions of QAL, for two groups and more than two groups.

5. I have derived the most efficient estimating equation for the regression problem of the mean QAL and obtained an improved estimating equation which is more efficient than the simple weighted estimating equation.

## d. Reportable outcomes

1. I was an invited speaker for the Joint Statistical Meetings, August 3-7, 2003, San Francisco, CA. My talk was titled "Statistical Inference for Quality-Adjusted Survival Time".

2. I gave a short course entitled "Statistical Inference of Quality Adjusted Lifetime" in International Chinese Statistical Association (ICSA) 2005 Applied Statistics Symposium at Washington, DC on Saturday, June 19, 2005.

3. Pandya, K.J., Morrow, G.R., Roscoe, J.A., **Zhao, H.**, Hickok, J.T., Pajon, E., Sweeney, T.J., Banerjee, T.K., Flynn, P.J. "Gabapentin for hot flashes in 420 women with breast cancer: A randomized double-blind placebo controlled trial", 2005, *Lancet*, **366**, 818-824.

4. Hickok, J.T., Roscoe, J.A., Morrow, G.R., Bole, C.W., **Zhao, H.**, Hoelzer, K.L., Dakhil, S.R., Moore, T., Fitch, T.R. "Serotonin receptor antagonists are no better than prochlorperazine for control of delayed nausea (DN) caused by doxorubicin: A URCC CCOP randomised study of 691 patients", 2005, *Lancet Oncology*, **6(10)**: 765-772.

5. Wang, H. and **Zhao, H.** "Regression Analysis of Mean Quality-Adjusted Lifetime with Censored Data". Revision Submitted.

## e. Conclusions

I have benefitted tremendously from the support of this grant. With the guaranteed research time, I have studied thoroughly the problem of making inference on quality adjusted lifetimes. I believe the research problem I have been working on is of great importance to breast cancer studies.

## References

Bang, H. and Tsiatis, A. A. (2002). Median Regression with Censored Cost Data. *Biometrics* **58**, 643-649.

Gelber, R.D., Gelman, R.S. and Goldhirsch, A. (1989). A quality-of-life oriented endpoint for comparing therapies. *Biometrics* **45**, 781-795.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* **47**, 663-685.

Pepe, M.S. and Fleming, T.R. (1989). Weighted Kaplan-Meier statistics - A class of distance tests for censored survival data. *Biometrics* **45**(2):497–507.

Robins, J. M. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. *AIDS Epidemiology-Methodological Issues*, N. Jewell, K. Dietz, and V. Farewell (eds), 297-331. Boston: Birkhuser.

Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* **89**, 846-866.

Zhao, H., and Tsiatis, A.A. "A Consistent Estimator for the Distribution of Quality Adjusted Survival Time", 1997, *Biometrika*, **84**, 339-348.

Zhao, H., and Tsiatis, A.A. "Efficient Estimation of the Distribution of Quality Adjusted Survival Time", 1999, *Biometrics*, **55**, 1101-1107.

Zhao, H., and Tsiatis, A.A. "Testing Equality of Survival Functions of Quality Adjusted Lifetime", 2001, *Biometrics*, **57**, 861-867.

# Regression Analysis of Mean Quality-Adjusted Lifetime with Censored Data

**Hongkun Wang***

Division of Biostatistics and Epidemiology, Department of Public Health Sciences,
University of Virginia, Charlottesville, Virginia 22908, U.S.A.
*email:* hkwang@virginia.edu
*phone:* (434)924-8514 *fax:* (434)243-5787

**and**

**Hongwei Zhao**

Department of Biostatistics and Computational Biology, University of Rochester,
601 Elmwood Avenue, Box 630, Rochester, New York 14642, U.S.A.

## SUMMARY

In clinical trials of chronic diseases such as AIDS, cancer or cardiovascular diseases, the concept of quality-adjusted lifetime (QAL) has received more and more attention. In this paper we consider the problem of how the covariates affect the mean QAL when the data are subject to right censoring. We allow a very general form for the mean model as a function of covariates. Using the idea of inverse probability weighting, we first construct a simple weighted estimating equation for the parameters in our mean model. We then find the form of the most efficient estimating equation, which yields the most efficient estimator for the regression parameters. Since the most efficient estimator depends on the distribution of the health history processes, and thus cannot be estimated nonparametrically, we consider different approaches for improving the efficiency of the simple weighted estimating equation using observed data. The applicability of these methods is demonstrated by both simulation experiments and a data example from a breast cancer clinical trial study.

Key words: Counting process; Estimating equation; Martingale process; Quality of life; Survival analysis.

# 1 Introduction

In studies that evaluate new therapies for chronic diseases such as cancer, AIDS or cardiovascular diseases, extending overall survival time may not be the only goal. Improving patients' quality of life is also important. Quality-adjusted lifetime (QAL) is a measure which combines patients' quality of life and survival time together and provides a useful summary for evaluating the treatment effect.

Quality adjusted lifetime has been studied by Goldhirsch *et al.* (1989), Glasziou *et al.* (1990) and Gelber *et al.* (1995). In their work, a patient's health history was partitioned into different health states, e.g. toxicity state during cancer treatment, period of good health, and disease relapse state. Each state was assigned a utility coefficient, usually ranging from 0 (death) to 1 (good health). The QAL, also called Quality-Adjusted Time Without Symptoms and Toxicity (Q-TWiST), is defined as the linear combination of the utility coefficients and the times spent in each health state. In a more general setting, the QAL is simply defined as the integration of utilities over a subject's survival time.

In most clinical trials, patients enter the study over a period of time, and we cannot always observe the QAL for every patient due to loss to follow-up and study termination. The inference on QAL thus has to be made using censored data. Censoring poses a unique problem for making inference on QAL, since even though we are willing to assume that censoring is independent of the health history process, the censored QAL is often correlated with the potential uncensored QAL due to the induced informative censoring problem (Gelber *et al.,* 1989). For example, people with poor quality of life will accrue QAL slowly, and when they are censored, they will have small censored QAL as well. Much research has been done on estimating the mean QAL (e.g., Glasziou *et al.,* 1990; Gelber *et al.,* 1991; Zhao and Tsiatis, 2000), or the survival distribution of QAL (e.g., Zhao and Tsiatis, 1997; Zhao and Tsiatis, 1999; Van der Laan and Hubbard, 1999) from censored data. However, in a real application, if is often of our interest to know how covariates affect the mean QAL.

An example that illustrates the use of regression models for QAL came from clinical studies conducted by the International Breast Cancer Study Group (IBCSG). The IBCSG Trial V (Cole *et al.,* 1993) was a randomized clinical trial investigating two treatments for the node-positive breast cancer: short duration chemotherapy (one month) and long duration chemotherapy (six or seven months). One thousand two hundreds and twenty nine patients were enrolled in the study with 413 patients randomized to the short term chemotherapy and 816 patients randomized to the long term chemotherapy. The median follow-up for the study was seven years. Six covariates were recorded from each patient upon enrollment in the trial, which include age, treatment, tumor size, tumore grades (medium or high), and number of nodes involved. It was of interest to learn how patients' mean QAL might dependent on these prognostic factors.

Different approaches have been proposed for the regression problems of QAL. Cole *et al.* (1993) used a partitioned health state model and fitted Cox proportional hazards regression models for each transition time from the start of the study to the end of different health states. The mean QAL corresponding to a specific covariate value can be obtained by integrating the survival curves for that covariate value. With this approach, however, one cannot directly assess the covariate effects on mean QAL from the regression parameter estimates. In order to know how a covariate affect mean QAL, one has to plug in different values for this covariate while fixing other covariates at some population averages. Fine and Gelber (2001) proposed an accelerated life model for the distribution of survival and quality-adjusted survival time. However, their interests are the distribution of QAL, not the mean QAL. The regression method related to mean quality-adjusted lifetime data was mentioned in Bang and Tsiatis (2002), but no semi-parametric efficiency study was performed.

In this paper, we will study the problem of regressing the mean QAL on the covariates. We will investigate how to construct estimating equations for the regression parameters and how to obtain more efficient estimators by using the semi-parametric theory developed by Robins and Rotnitzky (1992). We will assume a mean model for QAL, but will not make any additional assumption on the underlying distribution of the health history process. Censoring is assumed to be independent of the health history process. In the discussion section, we will consider the situation when this condition is not met. Due to limited follow-up time, we only consider QAL accumulated up to a time limit $L$, where $L$ is determined by the availability of data. The rest of this paper is organized as follows. In Section 2, we describe the regression model and discuss methods for obtaining efficient estimators for the regression parameters. It is followed by the simulation experiments in Section 3. The breast cancer data example is analyzed in Section 4 and finally, some concluding remarks are given in Section 5.

# 2 Estimating Equations for Regression Parameters

## 2.1 The Regression model and Assumptions

For the $i$th individual in the study, let the health history process be represented by $\{V_i(t), t \geq 0 \;\; i = 1, \cdots, n\}$. Denote $V_i^H(t) = \{V_i(u) : u \leq t\}$, the health history up to time $t$. Let $T_i$ be the survival time and $q$ be a known utility function mapping $V_i(t)$ to the interval $[0, 1]$. $q$ is assumed to be known for our purpose. In the final section, we discuss how to handle the situation when $q$ is not known to us. The $i$th individual's quality-adjusted lifetime (QAL), denoted as $Q_i$, is equal to

$$Q_i = \int_0^{T_i} q\{V_i(t)\}dt.$$

Denote the $i$th individual's censoring time by $C_i$. Censoring is assumed to be independent of the health history process $V_i(.)$. The distribution of $C$ is assumed to be continuous and is denoted as $K(t) = \Pr(C > t)$. Because of censoring, we can not make inference on QAL over the entire health history; we only consider the QAL accumulated within a time limit $L$. Consequently, the survival time of an individual will be truncated at $L$, that is, $T^L = \min(T, L)$. For ease of notation, we still use $T$ instead of $T^L$. We assume that $\Pr(C \geq L) > 0$.

Let $Z_i$ denote the $(p + 1) \times 1$ vector of covariates associated with the $i$th individual, with the first covariate being the constant 1. The observed data for $n$ individuals are the independently and identically distributed random quantities: $[X_i = \min(T_i, C_i), \Delta_i = I(T_i \leq C_i), V_i(t), 0 \leq t \leq X_i, Q_i(u) = \int_0^u q\{V_i(t)\}dt, u \leq X_i, Z_i, i = 1, \cdots, n]$. According to this definition, $Q_i = Q_i(\min\{T_i, L)\}$.

We assume that the mean QAL depends on the covariates in a very general form:

$$E(Q_i|Z_i) = g(\beta, Z_i), \tag{2.1}$$

where $\beta$ is a $(p+1) \times 1$ vector of parameters of interest. Special cases include $g(\beta, Z_i) = \beta' Z_i$, a linear regression model and $g(\beta, Z_i) = g(\beta' Z_i)$, a generalized linear regression model. Our goal is to make inference about $\beta$ in the mean model (2.1) for some pre-specified function $g$ from the observed censored quality of life and survival data.

## 2.2 Simple Weighted Estimating Equation

If complete data are observed, a consistent estimator $\hat{\beta}$ for $\beta$ in the mean model (2.1) can be obtained from the following estimating equation:

$$S_n^F(\beta) = \sum_{i=1}^{n} h(Z_i)\{Q_i - g(\beta, Z_i)\} = 0, \tag{2.2}$$

where $h(Z_i)$ is $(p+1)$-dimensional vector of functions of $Z_i$, and the superscript $F$ represents models for full data. From the semi-parametric theory (Robins and Rotnitzky, 1992), we know that the most efficient estimating equation for complete data case is the one with

$$h_{eff}^F(Z_i) = \mathrm{Var}(Q_i|Z_i)^{-1} \frac{\partial g(\beta, Z_i)}{\partial \beta}|_{\beta_{\mathbf{0}}},$$

where $\beta_{\mathbf{0}}$ is the true value of the parameters.

In the special case of a linear model where $g(\beta, Z_i) = \beta' Z_i$ and $\mathrm{Var}(Q_i|Z_i)$ is assumed to be a constant, the most efficient estimating equation is obtained by setting $h_{eff}^F(Z_i) = Z_i$, and hence

$$S_{n,eff}^F(\beta) = \sum_{i=1}^{n} Z_i(Q_i - \beta' Z_i) = 0.$$

This equation is the same as the ordinary least squares estimating equation for the linear regression models.

When censoring is present, $Q_i$ cannot be observed for everybody so the estimating equation (2.2) cannot be used. However, using the idea of inverse probability weighting, which was originally proposed by Horvitz and Thompson (1952), we can construct a simple weighted estimating equation for $\beta$ in our mean model (2.1) with censored data:

$$S_n(\beta) = \sum_{i=1}^{n} \frac{\Delta_i}{K(T_i)} h(Z_i)\{Q_i - g(\beta, Z_i)\} = 0,$$

where $\Delta_i = I(T_i \leq C_i)$, $K(T_i)$ is the survival probability for the censoring variable $C$ at time $T_i$. The consistency of the simple weighted estimating equation is shown by

$$\begin{aligned}
\mathrm{E}\{S_n(\beta)\} &= \mathrm{E}\left[\sum_{i=1}^{n} \mathrm{E}[\frac{\Delta_i}{K(T_i)} h(Z_i)\{Q_i - g(\beta, Z_i)\}|V_i^H(.), Z_i]\right] \\
&= \mathrm{E}\left[\sum_{i=1}^{n} \frac{h(Z_i)}{K(T_i)}\{Q_i - g(\beta, Z_i)\}\mathrm{E}\left\{I(C_i \geq T_i)|V_i^H(.), Z_i\right\}\right] \\
&= \mathrm{E}\left[\sum_{i=1}^{n} h(Z_i)\mathrm{E}[\{Q_i - g(\beta, Z_i)\}|Z_i]\right] \\
&= 0
\end{aligned}$$

Since $K(T_i)$ is not known to us, we can estimate it using the Kaplan-Meier estimator $\widehat{K}(T_i)$ (Kaplan and Meier, 1958). Hence, our simple weighted estimating equation becomes

$$S_n^{WT}(\beta) = \sum_{i=1}^{n} \frac{\Delta_i}{\hat{K}(T_i)} h(Z_i)\{Q_i - g(\beta, Z_i)\} = 0. \tag{2.3}$$

In the special case when $g(\beta, Z_i) = \beta' Z_i$ and $h(Z_i) = Z_i$, the estimating equation (2.3) has a closed-form solution for $\beta$ given by

$$\hat{\beta}^{WT} = \left\{\sum_{i=1}^{n} \frac{\Delta_i}{\hat{K}(T_i)} Z_i^{\otimes 2}\right\}^{-1} \left\{\sum_{i=1}^{n} \frac{\Delta_i}{\hat{K}(T_i)} Q_i Z_i\right\},$$

where we use the notation $a^{\otimes 2} = aa'$, $a^{\otimes} b = ab'$, for vectors $a$ and $b$.

The simple weighted estimator is easy to calculate, however, only the QAL for the subjects with observed failures are used in the estimating equation. Therefore, it cannot be efficient.

## 2.3 Efficiency Study

To develop the theory on efficiency study of the estimating equations, we use the counting processes and the associated martingale theory as described by Fleming and Harrington (1991). Let the filtration $\mathcal{F}(u)$ be the increasing sequence of $\sigma-$algebras generated by

$$\sigma\{I(C_i \leq t), t \leq u; I(T_i \leq x), V_i^H(x), 0 \leq x < \infty, Z_i, i = 1, \ldots, n\}.$$

We consider the martingale process $\mathcal{M}_i^C(u) = N_i^C(u) - \int_0^u \lambda^C(t) Y_i(t) dt$, where $N_i^C(u) = I(X_i \leq u, \Delta_i = 0)$, $Y_i(t) = I(X_i \geq t)$, $\lambda^C(t) = \lim_{h \to 0} \frac{1}{h} \Pr(t < C < t + h | C \geq t, T \geq t)$ is the hazard function for the censoring distribution.

From the semi-parametric theory for missing data processes developed by Robins and Rotnitzky (1992) and Robins et al. (1994), the influence function for the estimating equation for any regular asymptotic linear (RAL) estimators of $\beta_{\mathbf{0}}$ can be written as

$$D_i^h - \int_0^\infty \{D_i^h - G(D^h, u)\} \frac{d\mathcal{M}_i^C(u)}{K(u)} + \int_0^\infty \left[e\{V_i^H(u)\} - G[e\{V^H(u)\}, u]\right] \frac{d\mathcal{M}_i^C(u)}{K(u)} \qquad (2.4)$$

where $D_i^h \equiv h(Z_i)\{Q_i - g(\beta, Z_i)\}$ is the influence function for the complete data, $G(W, u) = E\{W_i I(T_i \geq u)\}/S(u)$ for any random variable or functional $W$, $S(u) = \Pr(T > u)$, and $e\{V_i^H(u)\}$ is any $(p+1)$-dimensional vector of functionals of the health history $V_i^H(u)$. It should be noted that the influence function for the simple weighted estimating equation (2.3) is simply the first two terms of (2.4) (Zhao and Tsiatis, 1997, equation A.7).

From Robins and Rotnitzky (1992), the most efficient estimating equation is obtained by choosing

$$e_{eff}\{V_i^H(u)\} = \mathrm{E}\{D_i^{h_{eff}}|V_i^H(u)\} = h_{eff}(Z_i)\mathrm{E}\{D_i|V_i^H(u)\},$$

where $D_i \equiv Q_i - g(\beta, Z_i)$, and

$$h_{eff}(Z_i) \quad = \quad \{\mathrm{Var}(Q_i|Z_i) + P(Z_i)\}^{-1}\frac{\partial}{\partial\beta}g(\beta, Z_i)|_{\beta_0},$$

with

$$P(Z_i) = \mathrm{E}\left[\int_0^\infty \frac{dN_i^C(u)}{K(u)}\mathrm{Var}\{D_i|V_i^H(u), Y_i(u) = 1, Z_i\}|Z_i\right] \qquad (2.5)$$

From the above results, the most efficient estimating equation can be formed by

$$S_{n,eff}(\beta) = \sum_{i=1}^n \frac{\Delta_i D_i h_{eff}(Z_i)}{\hat{K}(T_i)} + \sum_{i=1}^n \int_0^\infty \{e_{eff} - \hat{G}^\star(e_{eff}, u)\}\frac{dN_i^C(u)}{\hat{K}(u)} = 0, \quad (2.6)$$

where $\hat{G}^\star(W, u) = \sum_{i=1}^n W_i Y_i(u)/Y(u)$ is a consistent estimator for $G(W, u)$, for any functional W.

In theory, the asymptotic variance of $\hat{\beta}$ from solving (2.6) should achieve the semi-parametric efficiency bound, which means that $\hat{\beta}$ from (2.6) has the smallest variance among the class of all regular asymptotically linear estimators. However, it is not useful to use (2.6) for data analysis, since $e_{eff}$ and $h_{eff}$ depend on the unknown true population parameters which are difficult to estimate nonparametrically.

## 2.4 Improved Estimating Equation

Due to the difficulty in obtaining the most efficient estimating equation nonparametrically, we wish to find an estimating equation which can be obtained from the observed data, and which can be more efficient than the simple weighted estimating equation for any choice of $h(Z)$. In the subsequent section, we will discuss the issues of choosing $h(Z)$.

### 2.4.1 The Best-Coefficient Approach

We first consider a method for obtaining the improved estimating equation, which is similar to the approach appeared in Zhao and Tsiatis (1999) for obtaining improved estimators of mean QAL, and in Bang and Tsiatis (2002) for median regression of medical costs. We will call this approach the Best-Coefficient approach and denote it as BC.

For any chosen $e\{V_i^H(u)\}$, if we multiply the third term in (2.4) by a constant $\Gamma$

$$D_i^h - \int_0^\infty \{D_i^h - G(D^h, u)\} \frac{d\mathcal{M}_i^C(u)}{K(u)} + \Gamma \int_0^\infty \left[ e\{V_i^H(u)\} - G[e\{V^H(u), u\}] \right] \frac{d\mathcal{M}_i^C(u)}{K(u)} \tag{2.7}$$

where $\Gamma$ is equal to $\text{Cov}(W_1, W_2)\text{Var}(W_2)^{-1}$, with $W_1$ and $W_2$ being the second and third terms in (2.4), i.e.

$$W_1 = \int_0^\infty \{D_i^h - G(D^h, u)\} \frac{d\mathcal{M}_i^C(u)}{K(u)},$$

$$W_2 = \int_0^\infty \left[ e\{V_i^H(u)\} - G[e\{V^H(u), u\}] \right] \frac{d\mathcal{M}_i^C(u)}{K(u)},$$

then the variance of this influence function (2.7) will always be smaller than that of the simple weighted estimating equation. In practice, $\Gamma$ is not known, so it has to be estimated from available data, which will result in some additional variability for finite sample sizes. We will examine its finite sample performance in our simulation study.

We can derive an explicit formula for the BC estimator in the special case when $g(\beta, Z) = \beta' Z$ and $h(Z) = Z$. If we choose $e\{V_i^H(u)\} = Q_i(u)$, for example, we can get

$$\hat{\beta}^{BC} = A_1^{-1} A_2, \tag{2.8}$$

where

$$\begin{aligned}
A_1 &= \sum_{i=1}^{n} \frac{\Delta_i}{\hat{K}(T_i)} Z_i^{\otimes 2} + \hat{W}_2 [\hat{J}\{Q(u) \cdot Q(u)\}]^{-1} \hat{J}\{Z^{\otimes 2} \cdot Q(u)\}, \\
A_2 &= \sum_{i=1}^{n} \frac{\Delta_i}{\hat{K}(T_i)} Q_i Z_i + \hat{W}_2 [\hat{J}\{Q(u) \cdot Q(u)\}]^{-1} \hat{J}\{ZQ \cdot Q(u)\}, \\
\hat{W}_2 &= \sum_{i=1}^{n} \int_0^L \frac{dN_i^C(u)}{\hat{K}(u)} [Q_i(u) - \hat{G}^\star\{Q(u), u\}], \\
\hat{J}(X \cdot Y) &= \frac{1}{n} \int_0^L \{\hat{G}^\star(X \cdot Y, u) - \hat{G}^\star(X, u) \hat{G}^\star(Y, u)\} \frac{dN^C(u)}{\hat{K}(u)^2}, \qquad (2.9)
\end{aligned}$$

for any functionals $X$ and $Y$. If $X$ or $Y$ involves $Q_i$, then $\hat{G}^\star$ will be replaced by the $\hat{G}$ function

$$\hat{G}(X, u) = \frac{1}{n} \frac{1}{\hat{S}(u)} \sum_{i=1}^{n} \frac{\Delta_i X_i I(T_i \geq u)}{\hat{K}(T_i)}.$$

The detail of the derivation is given in the Appendix.

### 2.4.2 The Improved Estimating Equation with $\Gamma = 1$ and $Q_i(u)$ in Place of $\mathbf{E}\{Q_i | V_i^H(u)\}$

Our second strategy is to choose $\Gamma = 1$, and use $Q_i(u)$ in place of $\mathrm{E}\{Q_i | V_i^H(u)\}$ in the formula for $e_{eff}\{V_i^H(u)\}$, i.e. we choose

$$e\{V_i^H(u)\} \quad = \quad h(Z_i)\{Q_i(u) - g(\beta, Z_i)\} \equiv D_i^h(u).$$

The corresponding estimating equation, named the improved estimating equation and denoted as IMP, has the following form:

$$S_n^{IMP}(\beta) = \sum_{i=1}^{n} \frac{\Delta_i}{\hat{K}(T_i)} D_i^h + \sum_{i=1}^{n} \int_0^\infty \frac{dN_i^C(u)}{\hat{K}(u)} [D_i^h(u) - \hat{G}^\star\{D^h(u), u\}] = 0. \tag{2.10}$$

This estimating equation (2.10) is not guaranteed to be always more efficient than the simple weighted estimating equation (2.3). However, due to the usual correlation between $Q_i(u)$ and $\mathrm{E}\{Q_i|V_i^H(u)\}$, we expect this estimator to perform well in most realistic settings.

In the case when $g(\beta, Z) = \beta' Z$ and $h(Z) = Z$, this improved estimator has an explicit form $\hat{\beta}^{IMP} = \mathbf{C}_1^{-1}\mathbf{C}_2$, where

$$\begin{aligned}
\mathbf{C}_1 &= \sum_{i=1}^{n} \left\{ \frac{\Delta_i}{\hat{K}(T_i)} + \int_0^L \frac{dN_i^C(u)}{\hat{K}(u)} \right\} Z_i^{\otimes 2} - \sum_{i=1}^{n} \int_0^L \frac{\sum_{j=1}^n Y_j(u) Z_j^{\otimes 2}}{\hat{K}(u) Y(u)} dN_i^C(u), \\
\mathbf{C}_2 &= \sum_{i=1}^{n} \left\{ \frac{\Delta_i Q_i}{\hat{K}(T_i)} + \int_0^L \frac{Q_i(u)}{\hat{K}(u)} dN_i^C(u) \right\} Z_i - \sum_{i=1}^{n} \int_0^L \frac{\sum_{j=1}^n Q_j(u) Y_j(u) Z_j}{\hat{K}(u) Y(u)} dN_i^C(u).
\end{aligned}$$

### 2.4.3 The Estimating Equation Using Regression Approach

Our last strategy for improving efficiency is to choose $\Gamma = 1$ and estimate $\mathrm{E}\{Q_i|V_i^H(u)\}$ by regressing $\frac{\Delta_i Q_i K(u)}{K(T_i)}$ on observed covariates, using only those observations with $X_i \geq u$ (Robins and Rotnitzky, 1992). The resulting estimating equation will be in the same form as (2.10) except that $Q_i(u)$ is replaced by the estimate of $\mathrm{E}\{Q_i|V_i^H(u)\}$ from the regression approach.

To implement this idea, we may choose a linear regression model (LRG), regressing $\frac{\Delta_i Q_i K(u)}{K(T_i)}$ on some covariates that are predictive of QAL; or we may use a generalized additive model (GAM) (Hastie and Tibshirani, 1990; Van der Laan and Hubbard, 1999), which accommodates a nonparametric regression of $\frac{\Delta_i Q_i K(u)}{K(T_i)}$ on some functions of the health history process $V_i^H(u)$, e.g. $Q_i(u)$. We will compare the performance of these choices in our simulation studies.

## 2.5 Choice of $h(Z)$

Compared to the best choice of $h(Z)$ for the complete data case

$$h_{eff}^F(Z_i) = \mathrm{Var}(Q_i|Z_i)^{-1} \frac{\partial g(\beta, Z_i)}{\partial \beta}|_{\beta_0},$$

the optimum choice of $h(Z)$ for the incomplete data case is

$$h_{eff}(Z_i) = \{\mathrm{Var}(Q_i|Z_i) + P(Z_i)\}^{-1}\frac{\partial}{\partial\beta}g(\beta, Z_i)|_{\beta_0},$$

where $P(Z_i)$ is defined as (2.5). It is equivalent to down-weight the influence of $\mathrm{Var}(Q_i|Z_i)$, due to the added uncertainty about the variance of $Q_i$ given $Z_i$ for the censored observation. Since it is harder to estimate the second moment than the first moment given the high dimensional health history process, and secondly, using an incorrect model could potentially increase the variability of the estimating equation, we choose not to attempt to estimate $P(Z_i)$ and use instead the best choice of $h(Z)$ for the complete data case.

## 2.6 Variance Estimators for Regression Parameters

In this section, we derive the variance estimators for the regression parameters in our various estimating equations. Suppose $\hat{\beta}$ is the solution to an estimating equation and $\beta_0$ is the true value of the parameters. From Taylor's expansion, we have

$$\mathrm{Var}\{n^{\frac{1}{2}}(\hat{\beta} - \beta_0)\} \quad = \quad I_0^{-1}I_1I_0^{-1} \qquad (2.11)$$

where

$$I_1 = \mathrm{Var}\{n^{-\frac{1}{2}}S_n(\beta_0)\},$$

$$I_0 = -\mathrm{lim}_{n\to\infty}n^{-1}\frac{\partial S_n(\beta_0)}{\partial\beta} = \mathrm{E}\{h(Z_i)\frac{\partial g(\beta_0, Z_i)}{\partial\beta}\}.$$

In the special case when $g(\beta, Z) = \beta' Z$ and $h(Z) = Z$, we have

$$I_0 = \mathrm{E}Z_i^{\otimes 2}.$$

Based on the general influence function (2.4), $I_1$ is equal to

$$\mathrm{Var}\{D_i^h(\beta_0)\}+\mathrm{Var}\{W_1(\beta_0)\}+\mathrm{Var}\{W_2(\beta_0)\}-\mathrm{E}\{W_1(\beta_0)\otimes W_2(\beta_0)\}-\mathrm{E}\{W_2(\beta_0)\otimes W_1(\beta_0)\},$$

where $D_i^h(\beta_0)$, $W_1(\beta_0)$, and $W_2(\beta_0)$ are defined similarly as $D_i^h$, $W_1$, and $W_2$, with true parameters $\beta_0$ plugged in.

Using derivations similar to those in the Appendix, we can show that for large $n$, $I_1$, the asymptotic variance of $n^{-\frac{1}{2}}S_n(\beta_0)$, can be estimated by

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n}\frac{\Delta_i}{\hat{K}(T_i)}\{D_i^h(\beta_0)\}^{\otimes 2} + \hat{J}\{D^h(\beta_0)\otimes D^h(\beta_0)\} + \hat{J}[e\{V_i^H(u)\}\otimes e\{V_i^H(u)\}] \\ - \;& \hat{J}[D^h(\beta_0)\otimes\{e\{V_i^H(u)\}] - \hat{J}[\{e\{V_i^H(u)\}\otimes D^h(\beta_0)] \end{aligned} \tag{2.12}$$

where

$$\hat{J}(X\otimes Y) = \frac{1}{n}\int_0^L \{\hat{G}^\star(X\otimes Y, u) - \hat{G}^\star(X,u)\otimes\hat{G}^\star(Y,u)\}\frac{dN^C(u)}{\hat{K}(u)^2},$$

for any vectors of functionals of $X$ and $Y$. Similar as the definition for $\hat{J}(X\cdot Y)$, $\hat{G}^\star$ will be replaced by the $\hat{G}$ function if $X$ or $Y$ involves the random variable $Q_i$.

The variance of the simple weighted estimating equation (2.3) is just the first two terms in (2.12). Due to the special coefficient used in the BC estimating equation (2.7), its variance can be easily shown to be

$$\begin{aligned} &\widehat{\mathrm{Var}}\{n^{-\frac{1}{2}}S_n^{WT}(\beta_0)\} \\ - \;& \hat{J}[D^h(\beta_0)\otimes\{e\{V_i^H(u)\}][\hat{J}[e\{V_i^H(u)\}\otimes e\{V_i^H(u)\}]]^{-1}\hat{J}[\{e\{V_i^H(u)\}\otimes D^h(\beta_0)]. \end{aligned}$$

# 3 Simulation Experiments

In this section we conduct some simulation experiments to evaluate our proposed estimating equations for the parameters in our regression models. Similar to the IBCSG Trial V example which is to be presented in the next section, we consider patients entering the study first experience toxicity for a certain time, then a period of good health (TWiST), then their disease relapse followed by death. We use TOX to represent the time from the treatment initiation to end of toxicity, TR the time from treatment initiation to disease relapse, and OS the time from treatment initiation to death. The quality adjusted lifetime is defined as:

$$\begin{aligned} Q &= q_{TOX} * TOX + TWiST + q_{REL} * REL \\ &= q_{TOX} * TOX + (TR - TOX) + q_{REL} * (OS - TR), \end{aligned}$$

where $q_{TOX}$ is the utility coefficient for TOX, $q_{REL}$ the utility coefficient for the REL (the period between disease relapse and death, The utility coefficient for TWiST is assumed to be 1, and $q_{TOX} = q_{REL} = q = 0.5$.

We generate 5,000 simulations, each consisting of two groups of censored health status data with sample sizes varying from 100 to 400 for each group. Two scenarios are considered here. In the first scenario, TOX is uniformly distributed on [0, $T_1$] for group 1 ($T_1$=60) and uniformly distributed on [0, $T_2$] for group 2 ($T_2$=80); TR is exponentially distributed with hazard $\lambda_1 = 1/130$ for group 1 and hazard $\lambda_2 = 1/90$ for group 2, and both are truncated at $L_1 = 81$. OS is exponentially distributed with hazard $\lambda_3 = 1/140$ and truncated at $L_2 = 100$ for both groups. The censoring variables for both groups are uniform on [70, 116] and are independent of TOX, TR and OS, which results in the amount of censoring to be about 35% for group 1 and 36% for group 2. For each group, if TR is greater than OS, we set TR=OS. Similarly, if TOX is greater than TR, we set TOX=TR. The true mean QAL for group k (k=1,2) is

$$\begin{aligned} &(1-q) * \{\frac{1}{(\lambda_k + \lambda_3)^2 T_k}(1 - e^{-(\lambda_k+\lambda_3)T_k}) - \frac{1}{\lambda_k + \lambda_3}(1 - e^{-(\lambda_k+\lambda_3)L_1})\} \\ &+q * \{\frac{1}{\lambda_3}(1 - e^{-\lambda_3 L_2})\}. \end{aligned}$$

Plugging in the parameter values, we can obtain that the true mean QAL is 47.91 for group 1 and 43.89 for group 2. Using a linear regression model with treatment as a covariate, and group 2 as the reference group, the intercept and slope parameters in our regression model are 43.89 and 4.02, respectively. In the second scenario, group 2 is generated the same way as group 1 in the first scenario, resulting in an intercept of 47.91 and a slope of 0.

We calculate the estimates for the intercept and slope, using the WT equation (2.3), the BC equation (2.8), the IMP equation (2.10), and the regression method. In the regression method, we consider 3 different approaches to estimate $\mathrm{E}\{Q_i|V_i^H(u)\}$: 1), using the sample average of $\frac{\Delta_i Q_i K(u)}{K(T_i)}$ only from observations with $X_i > u$, conditioning on the treatment at each time $u$ (denoted as AVE); 2), fitting a linear regression model for $\frac{\Delta_i Q_i K(u)}{K(T_i)}$ only from observations with $X_i > u$, combining all the censoring points and using treatment as the covariate (denoted as LRG); 3), fitting a generalized additive model for $\frac{\Delta_i Q_i K(u)}{K(T_i)}$ only from observations with $X_i > u$ by smoothing on the $Q_i(u)$ (denoted as SM).

Table 1 and Table 2 are results from the two simulation experiments, respectively. The sample standard errors (SSE), the estimated standard errors (ESE), and the sample coverage probabilities (CP) of the true parameters by the 95% confidence intervals of those estimators are given. We also calculate the estimates if we use the true $\mathrm{E}\{Q_i|V_i^H(u)\}$ (denoted as TrueE), since in the simulation we know the true distributions hence $\mathrm{E}\{Q_i|V_i^H(u)\}$ can be obtained. However, in practice, this estimator cannot be used since we do not know the true expectation of $Q_i$ given the health history process $V_i^H(u)$.

From the results of our simulation studies, we can see that the biases for all the estimators are rather small, indicating that all the estimating equations give consistent estimates of the regression coefficients. The empirical sample variances are very close to the estimated variances from formula (2.12). The estimators LRG and AVE have bigger sample standard errors than the simple weighted estimator.

Using the smoothing approach SM does not improve the efficiency. As expected from the theory, plugging in the true expectation gives us the most efficient estimator. The IMP estimator performs the best among all the estimators not using the true expectation. When the sample size is 100, the coverage probability for the BC estimator is not very accurate, but it improves considerable as sample size increases. The coverage probabilities for all other estimators are very close to 0.95 even for small sample sizes.

# 4 Application

In the IBCSG Trial V, each patient experienced in sequence three health states: TOX (toxicity), TWiST (perfect health), and REL (disease relapse). We illustrate our methods with the quality of life coefficients $q_{TOX} = q_{REL} = 0.5$ and the time limit $L = 84$ (months). The amount of censoring is 29.3%. Similar to Cole *et al.* (1993), we consider six covariates recorded from each patient upon enrollment in the study: treatment group (0= short duration, 1=long duration); tumor size (0=less than 2 cm, 1= at least 2 cm); logarithm of age; medium tumor grade (0=not medium grade, 1= medium grade); high tumor grade (0=not high grade, 1= high grade); number of nodes involved (0=fewer than 4, 1 =at least 4). As in Cole *et al.* (1993), 94 patients are removed from the data due to missing values for tumor grade.

We first considered a linear model with the six covariates and the interaction terms between treatment (treat) and the other covariates – tumor size (tsize), medium tumor grade (mgrade), high tumor grade (hgrade), and number of nodes involved (nodegrp). However, none of the interaction terms are significant thus they are excluded from our final model. We calculate the estimators using the WT estimating equation (2.3), the IMP estimating equation (2.10), the BC estimating equation (2.8) and other regression approaches for estimating $\mathrm{E}\{Q_i|V_i^H(u)\}$: the linear regression approach (LRG), the sample average approach conditioning on the treatment (AVE), and the generalized additive model with $Q_i(u)$ being the only regressor in the smoothing model (SM). Table 3 shows the results.

Concentrating on the covariates estimates and their standard errors first, we see that the AVE approach which estimates the expectation of $Q_i|V_i^H(u)$ by using the sample average for each treatment group at each censoring time does not perform well. The estimated standard error is bigger than all other approaches. The LRG approach, which is similar to AVE but combines all the censoring points together, performs slightly better. For some covariates SM approach produces a smaller standard error, but it is bigger for other covariates. BC estimator is consistently better than WT, but the improvement is not very big. The best performing estimator is the IMP estimator, similar as what we see in the simulation study. We have also considered two other generalized additive models: smoothing $Q_i(u)$ with both treatment and $Q_i(u)$ as the regressors in the model, and smoothing $Q_i(u)$ with all the six covariates and $Q_i(u)$ as the regressors in the model. We found out that adding more regressors in the smoothing method did not make much difference, so the results from those two models are not included in the Table 3.

Using the estimates from the IMP approach, we find that all six covariates are significant. A subject who is older, who has smaller tumor size, smaller number of nodes involved, lower tumor grade, and who is on the long duration arm, has a longer expected quality-adjusted lifetime. This finding agrees with the the description provided in the caption of Table 1 of Cole *et al.* (1993). Similar as

Cole *et al.* (1993), a sensitivity analysis can be carried out with varying values of $q_{TOX}$ and $q_{REL}$. For any fixed values for covariates, a treatment option can be chosen based on different quality of life utility values.

# 5 Conclusion

In this paper we have developed methods on how to estimate the covariate effects on the mean quality-adjusted lifetime, and how to obtain more efficient estimating equations. The theory developed by Robins and Rotnitzky (1992) provides the key to finding the form of the most efficient estimating equation, however, it is a function of the health history and cannot be estimated non-parametrically. We examined different approaches for obtaining consistent estimators for regression coefficients.

We have assumed that the quality of life coefficient $q$ is fixed in our methods. However, in a real application, $q$ is often not known and has to be estimated from QOL questionnaires. A lot of research has been devoted to this area. There are instruments developed which can translate health states into quality of life coefficients. In the cases similar to our example when the number of health states are limited, we can perform a sensitivity analysis and find out the treatment advantages for each set of utility values.

The simulation studies show that the best performing estimator is the IMP estimator using $Q_i(u)$ in place of $\mathrm{E}\{Q_i|V_i^H(u)\}$. The estimators using regression method (linear regression, or additive models) do not perform well. The BC estimator should always have smaller variance than the simple weighted estimator from the large sample theory, however, the improvement is not very big from our simulation studies.

We have assumed that censoring is independent of the health history process. If this assumption is not true, and censoring depends on some known covariates, we can accommodate this situation by fitting a Cox proportional hazards model to estimate the censoring distribution. If the Cox regression model is true, we can still get consistent estimators for the regression coefficients.

From Robins and Rotnitzky (1992), $h(Z)$ can be optimized to improve the efficiency of the estimating equation. However, optimizing $h$ involves estimating the second moments of $Q_i$ which will introduce some extra variability. How much more efficiency we may gain if we try to obtain the optimized $h$ will be a subject of future research.

## Acknowledgments

We are very thankful to Andrea Rotnitzky for helpful discussions on our paper. We are also very grateful to the International Breast Cancer Study Group (IBCSG) for making the IBCSG trial V data available to us. Hongwei Zhao was supported by Department of Defense Breast Cancer Research Program DAMD170210558.

## REFERENCES

Bang, H. and Tsiatis, A. A. (2002). Median Regression with Censored Cost Data. *Biometrics* **58**, 643-649.

Cole, B. F., Gelber, R. D. and Goldhirsch, A. (1993). Cox regression models for quality adjusted survival analysis. *Statistics in Medicine* **12**, 975-987.

Fine, J. P. and Gelber, R. D. (2001). Joint regression analysis of survival and quality-adjusted survival. *Biometrics* **57**, 376-382.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*, New York: Wiley.

Gelber, R. D., Gelman, R. S., and Goldhirsch, A. (1989). A Quality of life oriented endpoint for comparing therapies. *Biometrics* **45**, 781-795.

Gelber, R. D., Goldhirsch, A., and Cavalli, F. (1991). Quality-of-life-adjusted evaluation of a randomized trial comparing adjuvant therapies for operable breast cancer (for the international breast cancer study group). *Annals of Internal Medicine* **114**, 621-628.

Gelber, R. D., Cole, B. F., Gelber, S., and Goldhirsch, A. (1995). Comparing treatments using quality-adjusted survival: The Q-TWiST method. *The American Statistician* **49**, 161-169.

Glasziou, P. P., Simes, R. J., and Gelber, R. D. (1990). Quality adjusted survival analysis. *Statistics in Medicine* **9**, 1259-1276.

Goldhirsch, A., Gelber, R. D., Simes, R. J., Glasziou, P.P., and Coates, A. for the Ludwig Breast Cancer Study Group (1989). Costs and benefits of Adjuvant therapy in breast cancer: A quality-adjusted survival analysis. *Journal of Clinical Oncology* **7**, 36-44.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*, New York: Chapman and Hall.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* **47**, 663-685.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53**, 457-481.

Robins, J. M. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. *AIDS Epidemiology-Methodological Issues*, N. Jewell, K. Dietz, and V. Farewell (eds), 297-331. Boston: Birkhuser.

Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* **89**, 846-866.

Van der Laan, M. J. and Hubbard, A. (1999) Locally efficient estimator of the quality-adjusted lifetime distribution with right-censored data and covariates. *Biometrics* **55**, 530-536.

Zhao, H. and Tsiatis, A. A. (1997). A consistent estimator for the distribution of quality adjusted survival time. *Biometrika* **84**, 339-348.

Zhao, H. and Tsiatis, A. A. (1999). Efficient estimation of the distribution of quality-adjusted survival time.*Biometrics* **55**, 1101-1107.

Zhao, H. and Tsiatis, A. A. (2000). Estimating mean quality adjusted lifetime with censored data. *The Indian Journal of Statistics* **62**, Series B, 175-188.

## APPENDIX

### *The Derivation of* $\hat{\beta}^{BC}$

In the special case when $g(\beta, Z) = \beta' Z$ and $h(Z) = Z$, $e\{V_i^H(u)\} = Q_i(u)$, we can construct the following estimating equation based on the influence function (2.7):

$$S_n^{BC}(\beta) \;=\; S_n^{WT}(\beta) + \hat{\Gamma} \sum_{i=1}^{n} \int_0^L \frac{dN_i^C(u)}{\hat{K}(u)} [Q_i(u) - \hat{G}^{\star}\{Q(u), u\}], \qquad \text{(A.1)}$$

where $\hat{\Gamma}$ is a consistent estimator for $\Gamma = \text{Cov}(W_1, W_2)\text{Var}(W_2)^{-1}$ and can be obtained as follows.

Since

$$\text{Var}(W_2) = \text{E} \int_0^L [Q_i(u) - G\{Q(u), u\}]^2 Y_i(u) \frac{\lambda^C(u)}{K(u)^2} du,$$

it can be estimated consistently by

$$\begin{aligned} \widehat{\text{Var}}(W_2) &= \frac{1}{n} \int_0^L \sum_{i=1}^{n} [Q_i(u) - \hat{G}^{\star}\{Q(u), u\}]^2 Y_i(u) \frac{dN^C(u)}{\hat{K}(u)^2 Y(u)} \\ &= \frac{1}{n} \int_0^L [\hat{G}^{\star}\{Q(u)^2, u\} - \hat{G}^{\star}\{Q(u), u\}^2] \frac{dN^C(u)}{\hat{K}(u)^2} \\ &= \hat{J}\{Q(u) \cdot Q(u)\}, \end{aligned}$$

where $\hat{J}(X \cdot Y)$ for any random variables $X$ and $Y$ is defined in (2.9). Next,

$$\begin{aligned} \text{Cov}(W_1, W_2) &= \text{E} \int_0^L \{D_i^h - G(D^h, u)\}[Q_i(u) - G\{Q(u), u\}] I(T_i \geq u) \frac{\lambda^C(u)}{K(u)} du \\ &= \text{E} \int_0^L D_i^h [Q_i(u) - G\{Q(u), u\}] I(T_i \geq u) \frac{\lambda^C(u)}{K(u)} du \\ &= \int_0^L S(u)[G\{ZQQ(u), u\} - G\{Q(u), u\} G(ZQ, u)] \frac{\lambda^C(u)}{K(u)} du \\ &\quad - \int_0^L S(u)[G\{Z^{\otimes 2} Q(u), u\} - G(Z^{\otimes 2}, u) G\{Q(u), u\}] \frac{\lambda^C(u)}{K(u)} du \beta \end{aligned}$$

It can be estimated consistently by

$$
\begin{aligned}
\widehat{\text{Cov}}(W_1, W_2) &= \frac{1}{n}\int_0^L [\hat{G}\{ZQQ(u), u\} - \hat{G}(ZQ, u)\hat{G}^{\star}\{Q(u), u\}]\frac{dN^C(u)}{\hat{K}(u)^2} \\
&\quad -\frac{1}{n}\int_0^L [\hat{G}^{\star}\{Z^{\otimes 2}Q(u), u\} - \hat{G}^{\star}(Z^{\otimes 2}, u)\hat{G}^{\star}\{Q(u), u\}]\frac{dN^C(u)}{\hat{K}(u)^2}\beta \\
&= \hat{J}\{ZQ \cdot Q(u)\} - \hat{J}\{Z^{\otimes 2} \cdot Q(u)\}\beta
\end{aligned}
$$

Using these results, the estimating equation (A.1) can be written as

$$
S_n^{BC}(\beta) = S_n^{WT}(\beta) + \hat{W}_2 * \hat{J}\{Q(u) \cdot Q(u)\}^{-1} * [\hat{J}\{ZQ \cdot Q(u)\} - \hat{J}\{Z^{\otimes 2} \cdot Q(u)\}\beta],
$$

from which the BC estimator can be obtained easily and shown in (2.8)

**List of tables:**

Table 1: With treatment effect. Bias, sample standard errors (SSE), estimated standard errors (ESE), coverage probabilities for 95% confidence intervals (CP), for the intercept and slope for different estimators.

| Sample | | Intercept | | | | Slope | | | |
|---|---|---|---|---|---|---|---|---|---|
| Size | Estimator | Bias | SSE | ESE | CP | Bias | SSE | ESE | CP |
| 100 | WT | -0.052 | 2.942 | 2.959 | 0.947 | -0.044 | 4.767 | 4.719 | 0.939 |
| | IMP | 0.023 | 2.418 | 2.439 | 0.945 | 0.040 | 3.639 | 3.662 | 0.945 |
| | BC | -0.363 | 2.958 | 2.894 | 0.927 | -0.078 | 4.790 | 4.597 | 0.926 |
| | AVE | -0.038 | 4.237 | 4.226 | 0.953 | 0.091 | 7.898 | 7.849 | 0.947 |
| | LRG | 0.056 | 3.240 | 3.274 | 0.948 | -0.064 | 5.558 | 5.556 | 0.942 |
| | SM | 0.064 | 3.001 | 2.986 | 0.944 | -0.019 | 5.062 | 5.071 | 0.943 |
| | TrueE | 0.119 | 2.382 | 2.405 | 0.946 | -0.086 | 3.526 | 3.554 | 0.945 |
| 200 | WT | -0.046 | 2.094 | 2.091 | 0.950 | -0.041 | 3.321 | 3.334 | 0.950 |
| | IMP | -0.017 | 1.715 | 1.727 | 0.954 | -0.008 | 2.582 | 2.592 | 0.951 |
| | BC | -0.195 | 2.077 | 2.033 | 0.941 | -0.061 | 3.337 | 3.278 | 0.942 |
| | AVE | -0.017 | 2.844 | 2.912 | 0.953 | 0.039 | 5.329 | 5.386 | 0.952 |
| | LRG | 0.015 | 2.310 | 2.304 | 0.950 | -0.051 | 3.886 | 3.900 | 0.951 |
| | SM | 0.053 | 2.266 | 2.109 | 0.947 | 0.013 | 3.597 | 3.605 | 0.950 |
| | TrueE | 0.099 | 1.684 | 1.702 | 0.949 | -0.062 | 2.502 | 2.514 | 0.951 |
| 400 | WT | -0.020 | 1.502 | 1.497 | 0.952 | -0.035 | 2.441 | 2.435 | 0.954 |
| | IMP | 0.004 | 1.237 | 1.221 | 0.949 | -0.004 | 1.851 | 1.843 | 0.954 |
| | BC | -0.094 | 1.488 | 1.456 | 0.947 | -0.053 | 2.317 | 2.309 | 0.948 |
| | AVE | 0.003 | 2.045 | 2.045 | 0.950 | -0.030 | 3.760 | 3.775 | 0.947 |
| | LRG | 0.007 | 1.612 | 1.599 | 0.953 | -0.045 | 2.807 | 2.798 | 0.953 |
| | SM | -0.026 | 1.547 | 1.535 | 0.954 | 0.004 | 2.638 | 2.643 | 0.951 |
| | TrueE | 0.087 | 1.213 | 1.203 | 0.950 | -0.043 | 1.803 | 1.791 | 0.952 |

Table 2: No treatment effect. Bias, sample standard errors (SSE), estimated standard errors (ESE), coverage probabilities for 95% confidence intervals (CP), for the intercept and slope for different estimators.

| Sample |  |  | Intercept |  |  |  | Slope |  |  |
|---|---|---|---|---|---|---|---|---|---|
| Size | Estimator | Bias | SSE | ESE | CP | Bias | SSE | ESE | CP |
| 100 | WT | -0.067 | 3.217 | 3.192 | 0.947 | -0.071 | 4.974 | 4.906 | 0.944 |
|  | IMP | 0.019 | 2.649 | 2.648 | 0.948 | -0.064 | 3.797 | 3.808 | 0.946 |
|  | BC | -0.423 | 3.151 | 3.006 | 0.935 | -0.052 | 5.037 | 4.830 | 0.931 |
|  | AVE | 0.038 | 4.539 | 4.584 | 0.957 | -0.077 | 8.099 | 8.163 | 0.957 |
|  | LRG | 0.040 | 3.552 | 3.534 | 0.946 | 0.068 | 5.815 | 5.768 | 0.946 |
|  | SM | 0.092 | 3.617 | 3.217 | 0.941 | -0.055 | 5.890 | 5.252 | 0.945 |
|  | TrueE | 0.018 | 2.601 | 2.609 | 0.947 | -0.062 | 3.660 | 3.698 | 0.948 |
| 200 | WT | -0.041 | 2.274 | 2.260 | 0.946 | 0.014 | 3.537 | 3.464 | 0.946 |
|  | IMP | 0.015 | 1.867 | 1.875 | 0.950 | 0.007 | 2.752 | 2.693 | 0.945 |
|  | BC | -0.201 | 2.205 | 2.151 | 0.940 | 0.013 | 3.508 | 3.439 | 0.932 |
|  | AVE | 0.022 | 3.162 | 3.174 | 0.951 | 0.032 | 5.595 | 5.632 | 0.954 |
|  | LRG | 0.014 | 2.503 | 2.496 | 0.947 | 0.013 | 4.123 | 4.063 | 0.945 |
|  | SM | 0.013 | 2.287 | 2.276 | 0.952 | -0.010 | 3.749 | 3.703 | 0.947 |
|  | TrueE | 0.014 | 1.836 | 1.847 | 0.951 | 0.041 | 2.657 | 2.613 | 0.952 |
| 400 | WT | -0.013 | 1.584 | 1.598 | 0.949 | 0.007 | 2.412 | 2.405 | 0.946 |
|  | IMP | -0.006 | 1.333 | 1.327 | 0.952 | 0.004 | 1.896 | 1.906 | 0.950 |
|  | BC | -0.093 | 1.531 | 1.528 | 0.944 | 0.008 | 2.394 | 2.387 | 0.943 |
|  | AVE | -0.003 | 2.227 | 2.218 | 0.948 | 0.020 | 3.933 | 3.928 | 0.952 |
|  | LRG | 0.012 | 1.740 | 1.763 | 0.949 | -0.001 | 2.847 | 2.865 | 0.947 |
|  | SM | 0.009 | 1.674 | 1.669 | 0.951 | 0.007 | 2.639 | 2.644 | 0.952 |
|  | TrueE | -0.007 | 1.313 | 1.307 | 0.949 | 0.009 | 1.837 | 1.849 | 0.950 |

Table 3: Estimates for the regression coefficients (in months of quality-adjusted time) and their standard errors (ESE), for different estimators for the breast cancer example.

| Estimator | Intercept<br>(s.e.) | log(age)<br>(s.e.) | tsize<br>(s.e.) | nodegrp<br>(s.e.) | mgrade<br>(s.e.) | hgrade<br>(s.e.) | treat<br>(s.e.) |
|---|---|---|---|---|---|---|---|
| WT | 24.439 | 13.579 | -5.381 | -14.315 | -9.118 | -18.498 | 5.536 |
| | (15.546) | (3.831) | (1.873) | (1.711) | (2.241) | (2.411) | (1.750) |
| IMP | 23.436 | 13.532 | -4.499 | -14.372 | -7.824 | -17.489 | 4.977 |
| | (13.646) | (3.355) | (1.734) | (1.547) | (2.170) | (2.296) | (1.579) |
| BC | 25.079 | 13.486 | -5.439 | -14.395 | -9.137 | -18.570 | 5.469 |
| | (15.477) | (3.819) | (1.867) | (1.705) | (2.239) | (2.405) | (1.745) |
| AVE(trt) | 50.121 | 6.501 | -2.317 | -16.260 | -6.200 | -18.543 | 4.848 |
| | (16.663) | (4.097) | (2.274) | (1.994) | (3.196) | (3.253) | (2.027) |
| LRG | 22.976 | 14.072 | -5.351 | -14.078 | -9.829 | -18.955 | 5.453 |
| | (15.687) | (3.819) | (2.138) | (1.789) | (3.019) | (3.093) | (1.851) |
| SM(culU) | 17.598 | 15.035 | -4.657 | -13.880 | -8.702 | -17.636 | 5.212 |
| | (14.667) | (3.585) | (1.950) | (1.675) | (2.616) | (2.712) | (1.708) |